加微信：642945106 发送“ 赠送 ”领取赠送精品课程
发数字“ 2 ”获取众筹列表

07 | 性能数据的展示：一图胜千言，说出你的数据故事

2019-12-11 庄振运

讲述：庄振运
时长 14:05　大小 12.92M

你好，我是庄振运。

这一讲我们来探讨一下如何把性能数据合情合理地展示出来，让别人一目了然。如果你想有效地说服别人，那么你展示的过程就要像讲一个有趣的故事，娓娓道来，让别人爱听才行。

数据的展示根据场景有不同的目的，所以不能千篇一律，需要量体裁衣。每一种场景下，数据展示要根据你的**具体目的**、**听众的特点**和**内容的特点**而采用合适的图表。然后用这些图表做支持，把一个精美的数据分析的“故事”讲出来。

性能数据展示的目标和目的

我们先理清一下性能数据展示的目的是什么。

我们前面讲过，性能工作相对复杂，牵扯的模块和构件较多，而且有时候也要牵扯几个不同的部门。数据分析有时是为了性能优化，有时是为了根因分析等等。所以性能数据展示的目的也就有很多种，主要有如下三种：

1. 向上级报告性能趋势和流量预测的结果；
2. 向运维部门描述性能问题的根因分析；
3. 向开发部门建议性能提升和代码优化。

目标不同，听众不同，如何展示数据自然也就不能相同。但不管如何，都要做到有根有据，条理清楚，层次分明。这样大家才会被你和你展示的数据所说服。具体来讲，你希望你的听众：

1. 同意你的理论方法和过程；
2. 信服你的推理分析；
3. 理解问题的核心；
4. 看出问题的根因；
5. 同意你的建议和方案；

性能数据展示的挑战和难点

目的清楚了，下面就是怎么做了。但是我们必须认识到，相对于一般数据的展示，性能数据的展示有其独到的特点、挑战和难点：

数据量大：做性能测量时往往性能指标很多，每个指标一般都是一个时间序列。一个复杂系统 / 性能问题有几十甚至几百个性能指标是经常遇到的，所以总的数据样本经常以百万以上计。

数据复杂：几十或几百的性能指标互相交互和影响，需要考虑它们之间的复杂关系。

性能问题复杂：性能分析不是简单地死扣数据，还需要考虑互联网系统的协议，计算机的设计，和软硬知识的结合。

牵扯的模块多：很多性能问题都不是孤立的，都和其他模块和子系统，甚至客户请求有关系。

数据展示的经验

在考虑用什么方式来展示数据时，的确有些讲究，这就需要多借鉴别人的经验。在这方面，我总结了下面六条经验。

1. **按客人来备菜**：既然展示的听众不同，所以千篇一律的用同一种展示方法绝不可行。比如向领导和管理层汇报时，就要注重宏观层次，技术不要钻地太深。
2. **有啥菜吃啥饭**：要根据手头数据的特点来决定展示方法。比如你有好几组相关的数据，有很有趣的趋势，那就要用线图来展示趋势和相关性。
3. **给足上下文**：很多工程人员容易犯的错误，就是想当然地以为别人也都了解问题的背景，然后一上来就展示很多细节。其实除了你自己，没有第二个人更了解你要解决的问题和要展示的数据，所以一定要给足背景介绍和上下文信息。
4. **用图讲故事**：人人喜欢听故事，而且是有趣的故事。如果你能把整个分析和推理的过程，变成一个引人入胜和图文并茂的故事来讲，那我保证你的展示会非常地成功。
5. **和听众交互**：尽量鼓励听众参与讲故事的过程。有两种类型的数据叙事：叙述型和探索型。叙述型是通过描述告诉观众具体结论；而探索型是鼓励观众探索数据以得出结论。虽然探索型叙事稍微多花一点时间，但若能成功运用，听众更容易被说服；因为结论是他们自己得出的，而不是你告诉他们的。
6. **总结重要点**：在数据展示的最后，一定要简洁明了地总结你的展示。比如你希望听众最后只记住三句话，是哪三句呢？根据你的展示目的，这三句或是相关数字，或是趋势，或是问题本质，或是解决方案。

数据图表的种类

数据往往是枯燥的，所以我们要想办法变枯燥未有趣。用图表来展示数据就是一个好方法，但关键是要用合适的图表来展示不同的数据。

虽然数据图表的种类很多，但常用的就 10 种左右。我来简单地给你总结一下这几种常用的图表，分别适合什么情况和表示什么样的数据。

表格（Table）

防止断更 请务必加
首发微信：17161243665

表格你一定经常用，它的优点是可以结构化的方式显示大量信息；不幸的是，这个优点恰恰也是缺点。很多用户其实对数据的趋势比具体的数值更有兴趣，这种情况下用表格就不是那么直观。

线图（Line Chart)

如果我们希望针对一个变量，显示一段时间内这个变量的变化或趋势，线图就最合适了。比如我们经常表示的时间序列图就是最直白的线图。

另外，它也适用多个变量的情况，可以很直观地显示两个或多个变量之间的关系，比如趋势和相关性。

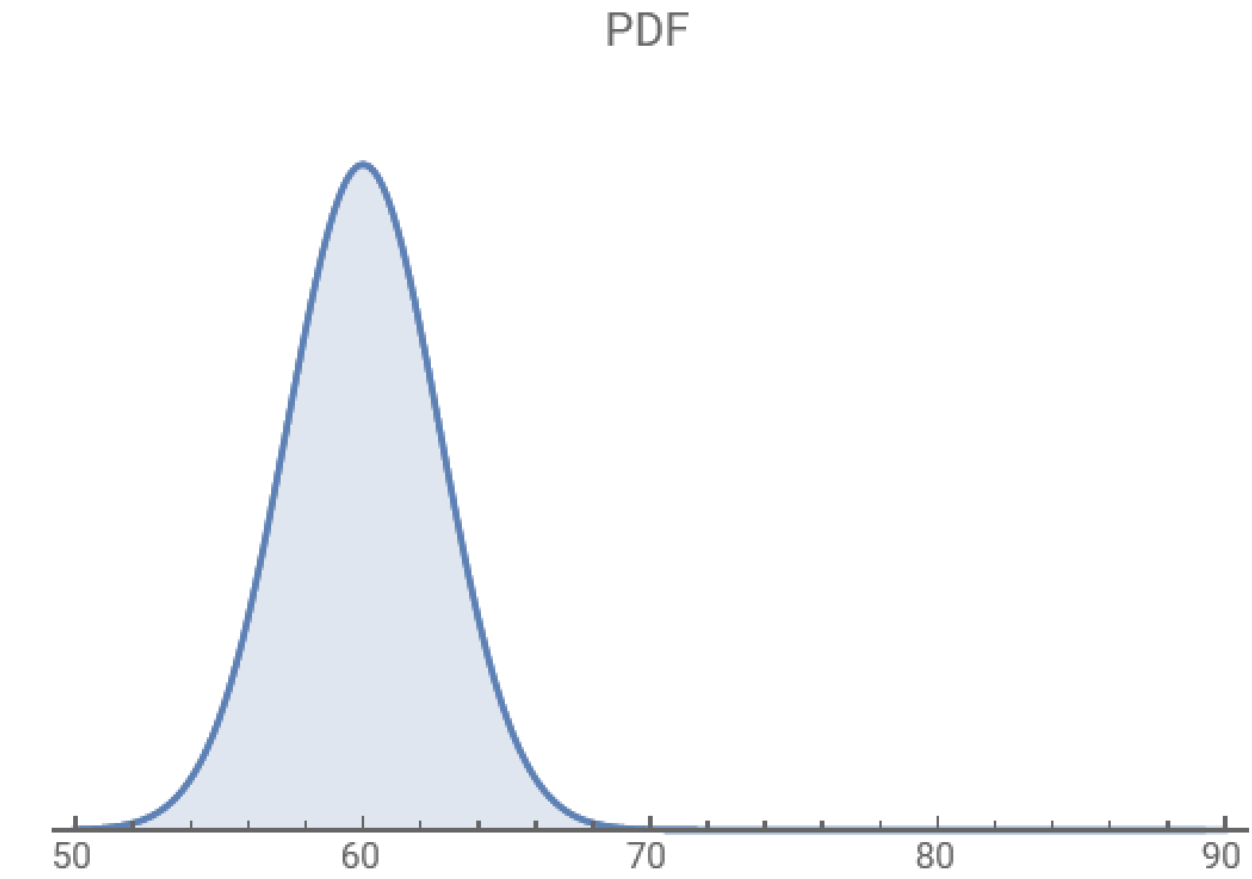
PDF 和 CDF 图

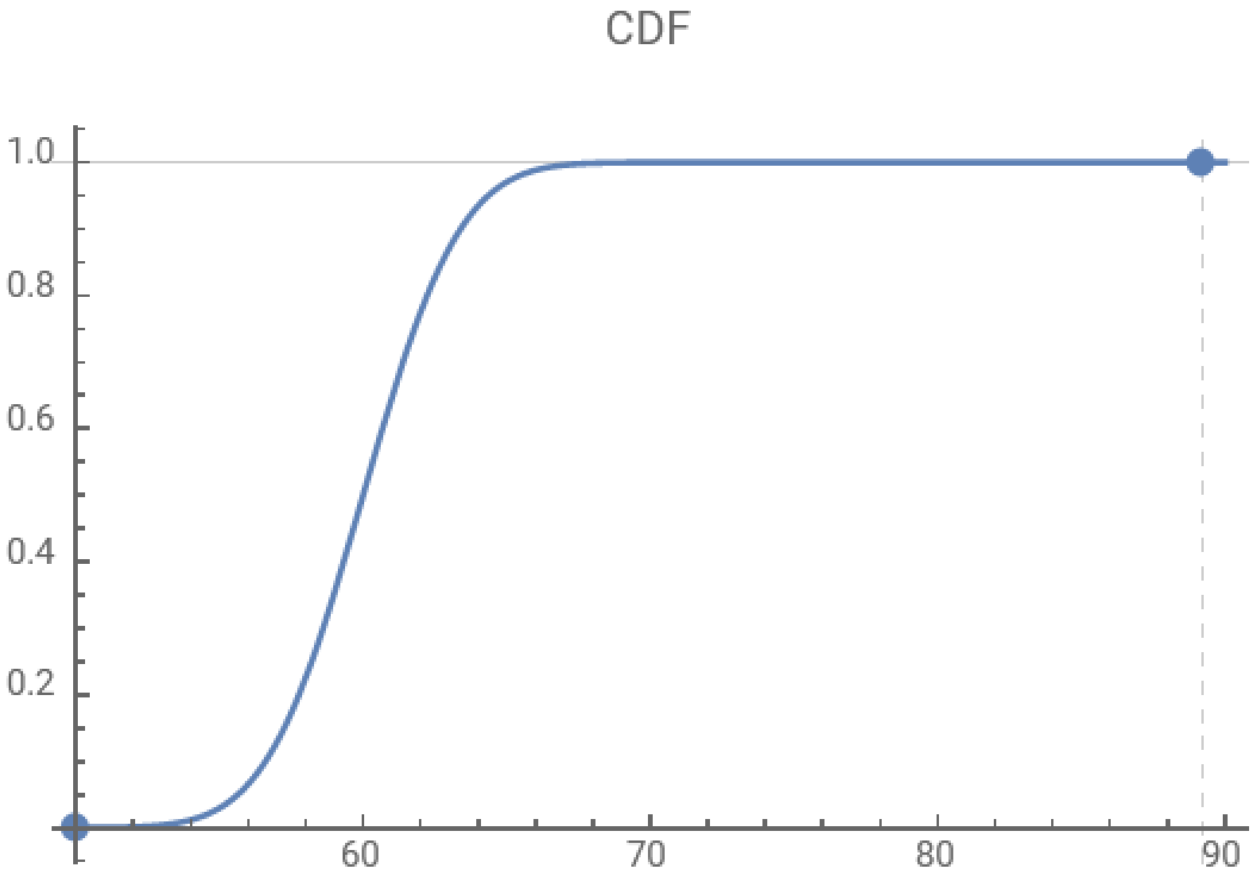
对于一个变量的概率数据，我们经常会显示这两个互相关联的图。二者可以看作是稍微特殊的线图。

PDF：概率密度函数（Probability Density Function）是连续型随机变量的概率密度函数（或简称为密度函数）。就是一个描述这个随机变量的输出值，在某个确定的取值点附近的可能性的函数。

CDF: 累积分布函数 (Cumulative Distribution Function)，又叫分布函数，是概率密度函数 PDF 的积分。换句话说，PDF 就是 CDF 的导数。CDF 能完整描述一个实随机变量 X 的概率分布。

我们举一个前面讲过的例子，比如标准正态分布。它的 PDF 和对应的 CDF 是这个样子。



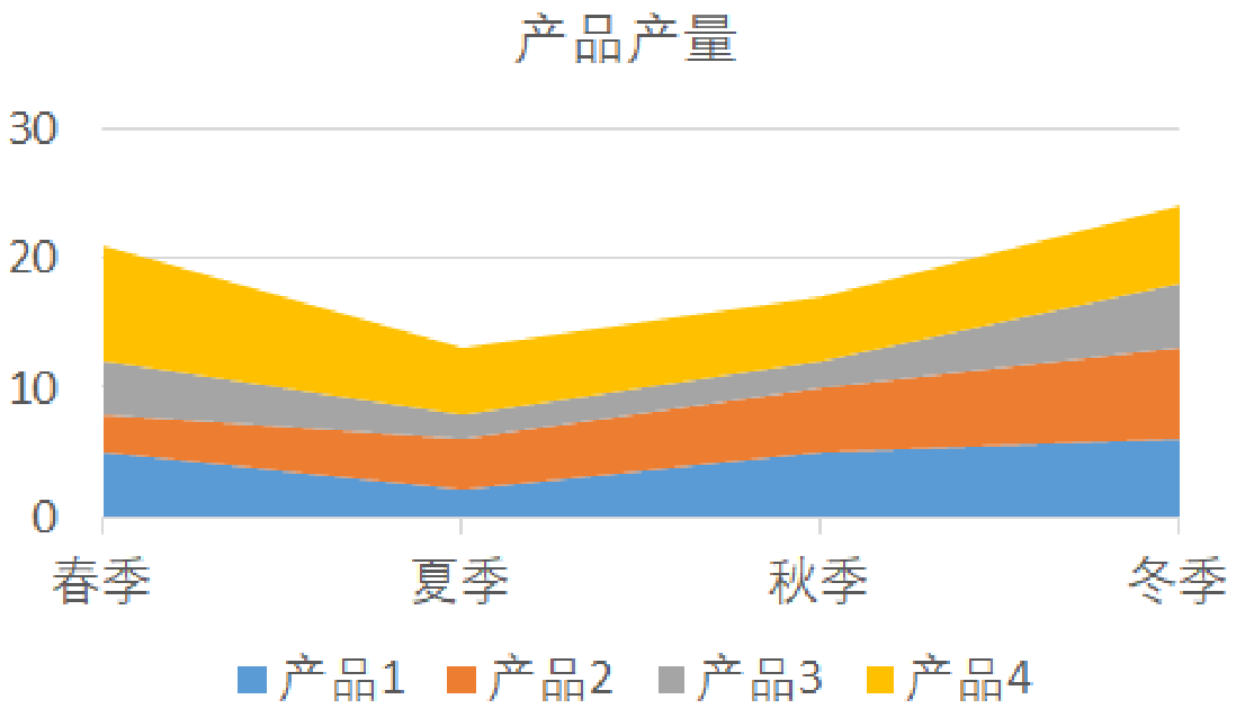


面积图（Area Charts）

面积图类似于线图，但两者之间有细微的差别。面积图的重点是阴影线下方的区域。

如果想用面积图表示几组数据，那么具体也有两种选择：可以选择采用叠加面积图或非叠加面积图。

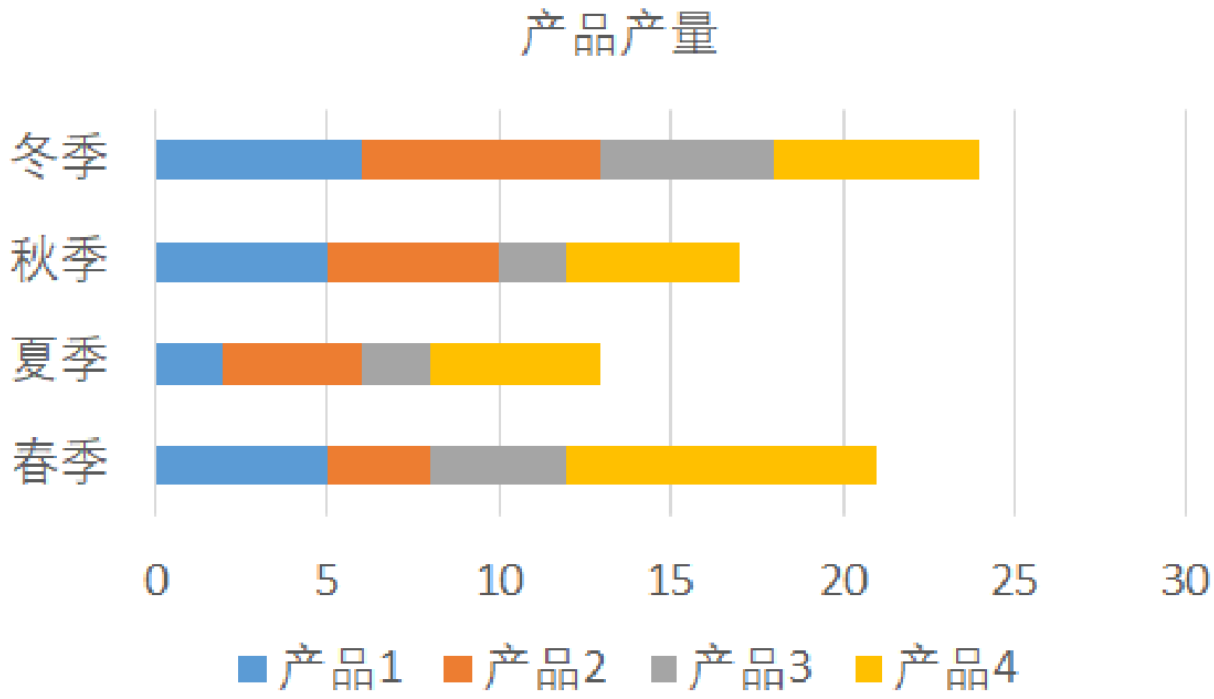
比如下图就是一个叠加面积图，它显示了几个产品在不同季度的单季产量和总产量。



柱状图和条形图（Bar Charts）

用于比较不同类别的数量，比较容易理解。如果你想表示好几个变量，那么也可以选择叠加方式来显示。

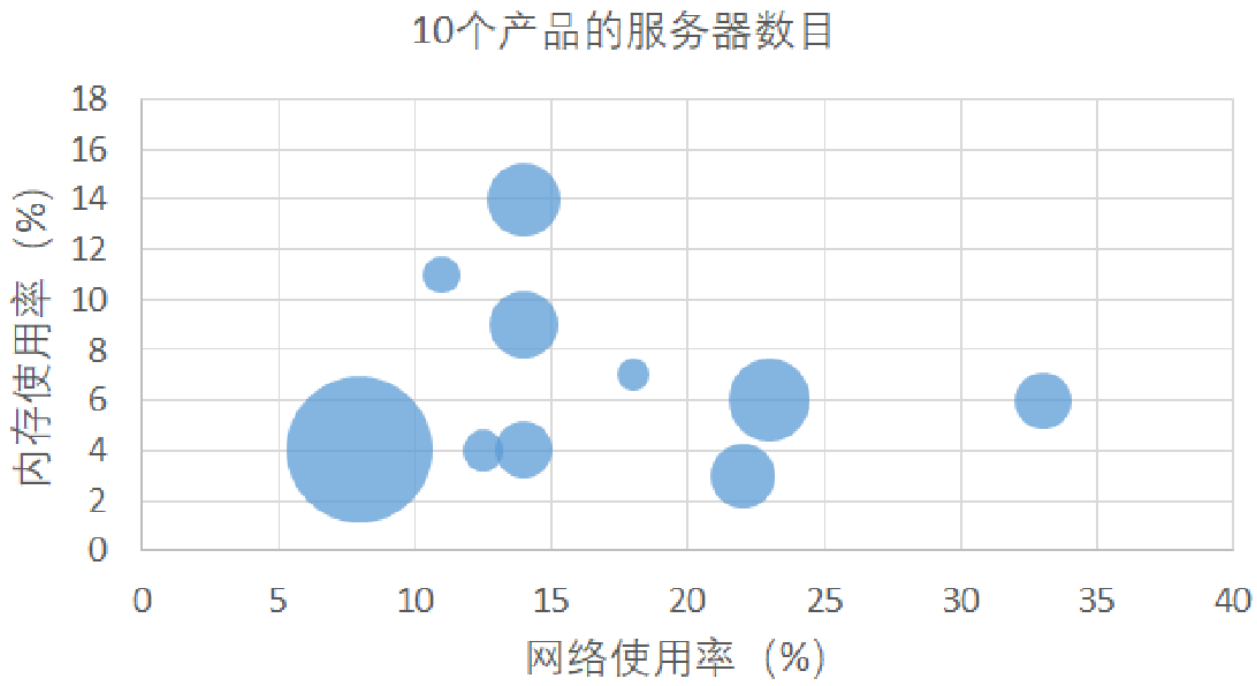
比如下面这个柱状叠加图，显示的是和上面的面积图同样的数据，只是表示的方法不同。



散点图（Scatter Plots）和气泡图（Bubble Charts）

散点图显示沿两个轴绘制的两个变量的值，用点的模式揭示它们之间存在的任何相关性。比如两个变量分别是 CPU 使用量和客户吞吐率，那么我们可以期望散点图会显示比较强的同一趋势。

气泡图类似于散点图，但它可以显示三个数据项之间的变化。除了两个变量分别为是横轴和纵轴外，气泡的大小代表第三个变量。



我们举一个例子。比如一个公司里面有十个互联网产品，每个产品运行在不同数目的服务器上面。这些产品分别有不同的内存使用率和网络使用率。假设我们想分析一下产品的内存和网络的使用情况，然后决定下一个季度购买什么样的服务器。我们就可以用气泡图来表示，网络和内存的使用率来作为横轴和竖轴，气泡的大小代表每个产品的服务器数目。

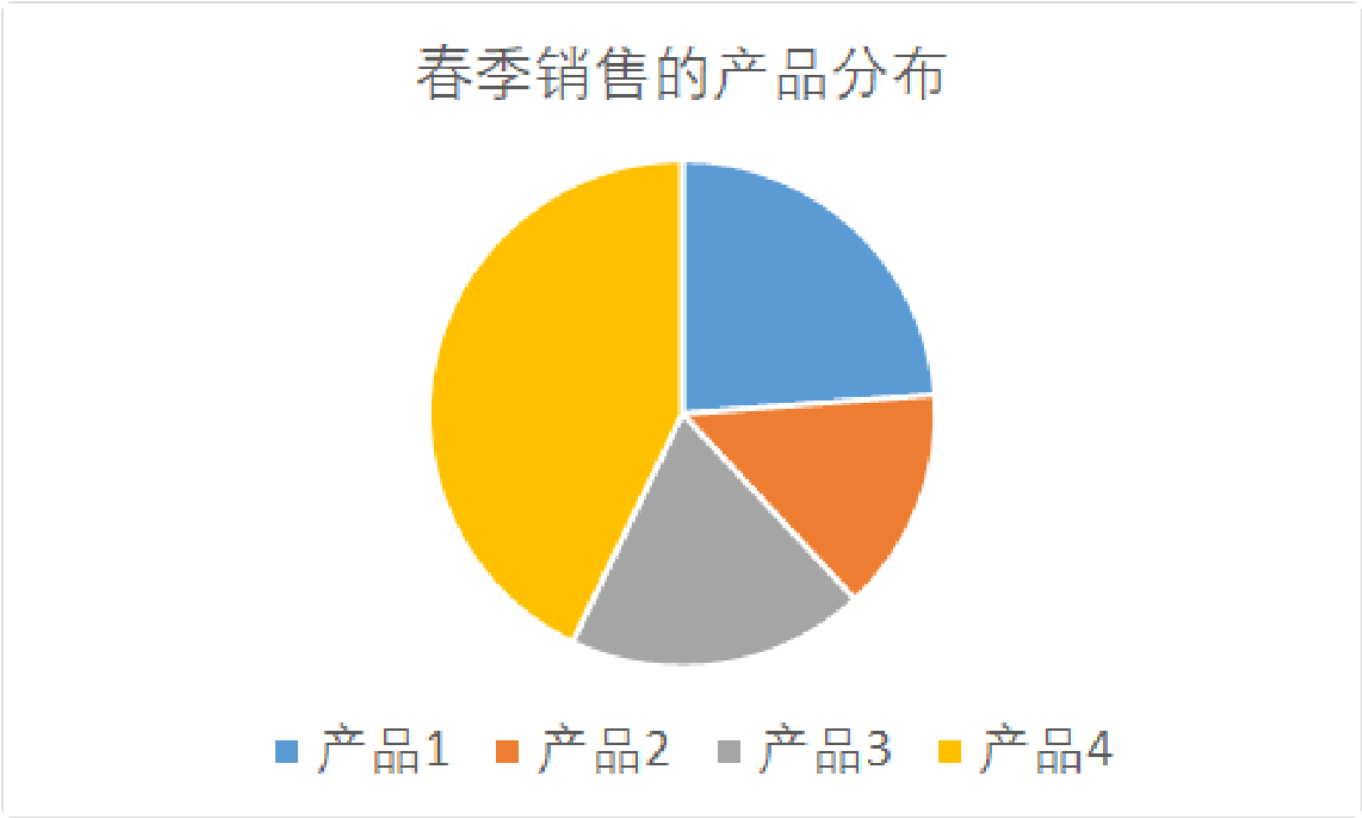
通过这个气泡图，我们非常直观地发现，某个产品的气泡比其他产品都大，而且它的两种资源使用率都很低。整体来言，下个季度新采购的服务器可以不需要那么多内存和网络资源。

饼图和圆环图（Pie Charts， Donut Charts）

当需要显示比例数据或者百分比时，饼图最佳。由于饼图表示零件与整个实体之间的大小关系，因此零件需要总和必须为有意义的整体。

一个适合饼图表示的性能数据例子是客户请求的来源分布。但你在使用时需要注意的是，饼图最好用来显示六个或更少的类别。如果太多类别的话反而描述不清楚。

举一个前面讲面积图时候用过的例子，公司的四种产品在春季的销售分布，就可以用下面这个饼图表示。



树形图（Treemaps）

树形图对于显示类别和子类别之间的层次结构和比较值非常有用，同时也能保留较多细节。树形图可以帮助我们很直觉地感知哪些区域最重要。另外，还可以通过将颜色编码的矩形嵌套在彼此内部，来更好地实现目的，并用加权以反映它们在整体中的份额。

比如下面这个树形图，它描绘了一个国外的产品，采用不同营销渠道的价值，然后按国家 / 地区细分。

这里的营销渠道包括 Email、Social Media、AdWords 等。从这个图中，你一眼就能看出从营销渠道上来看，AdWords 是最成功的营销渠道，因为它对应的面积最大。但是从国家层面来讲，美国（United States）是所有渠道中最有价值的目的地，同样是因为对应的面积最大。

热图（Heatmaps）

热表以表格格式来表示数据，其中每个格子是用颜色，而不仅仅是用数字来展示。这些颜色分别对应包含定义的范围，如绿色代表小的值，黄色代表一般的值和红色代表大的值。

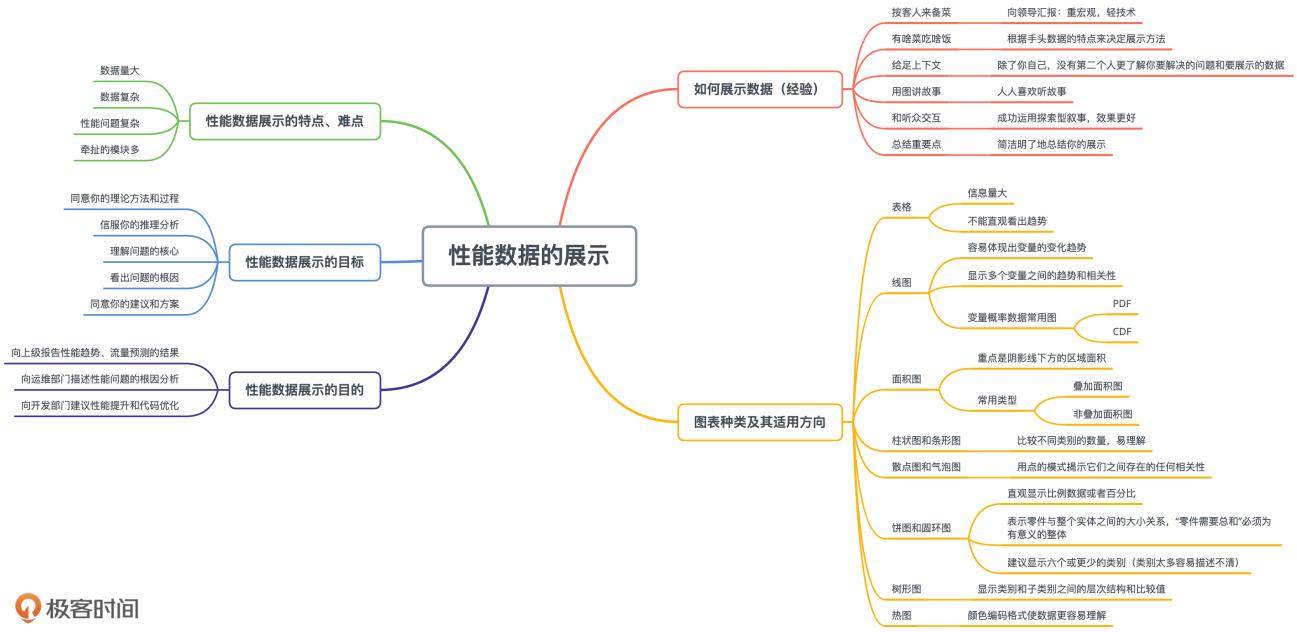
比如分析几台服务器的在一天里面 CPU 的使用率。我们可以把每台服务器在每个时段的 CPU 使用率分为小、较小、中等、较大和大等几个范围。然后相应地着色。

我们看下图，横坐标代表时段，纵坐标是服务器，每个值就是 CPU 使用率。

CPU使用率(%)	早上6点	早上8点	早上10点	中午12点	下午2点	下午4点	下午6点
服务器1	20	24	87	44	34	23	12
服务器2	32	56	45	32	61	55	19
服务器3	41	52	34	22	31	44	23
服务器4	25	34	23	34	36	34	12
服务器5	32	55	77	38	52	56	41

虽然每个使用率其实是数字，但是把数字编码为颜色的好处是，颜色编码格式使数据更容易理解。比如，我们马上就可以看出早上 10 点的时候，服务器 1 和 5 的 CPU 使用率很高，因为它们的格子是红色。

总结



我们都学过宋朝的词人柳永的词，他有两句词是这么写的：

“便纵有千种风情，更与何人说？”

这里他其实是在感叹自己虽有满腹经纶和深厚情感，但无人可以诉说。我们虽然不是词人，但在做过性能测试或者根因分析后，可能经常会有很多新奇的发现，希望向同事和领导讲解。

但是性能问题往往牵涉很广，描述起来并不容易，别人可能很难理解。这时候我们的感觉就是“我有千种风情”，可是如何说啊？

希望通过这一讲，能让你对在图文并茂地来“诉说”这方面能有些提升。

思考题

对本讲介绍的几种图形，你对哪些比较熟悉并经常使用？

哪些是不熟悉的但觉得可能会很有用的？

对不熟悉的图形，你能不能花十几分钟时间去仔细学习一下并尝试使用呢？

欢迎你在留言区分享自己的思考，与我和其他同学一起讨论，也欢迎你把文章分享给自己的朋友。

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇　06 | 性能数据的分析：如何从大量数据中看出想要的信号？

下一篇　08 | 经验总结：必须熟记的一组常用性能数字

精选留言 (7)

写留言

whenloce
2019-12-11

用的什么画图工具?

展开

作者回复: 我一般用Excel比较多，有时用Matlab。其实网上有些在线的画图网站也不错。

1

西西弗与卡夫卡
2019-12-11

经常使用的有表格和饼图。前者用来显示详细数据，后者用来显示那些要素占比大。

NonStatic
2019-12-11

我以前用Azure Monitor的query画过图，后来还用Python自己分析数据画。pyecharts是我用的挺开心的一个库，推荐给大家。

展开

许童童
2019-12-11

相比于页面中千篇一律的Table表格展示数据，使用这些可视化的数据图表确实可以更直观地反映出数据的重点，就好像数据会说话一样，同样的，在日常开发中，我们也可以在表格中加一些颜色，让数据更容易观测，热图：我觉得是一个很好的实践。

展开

Q
2019-12-11

用得最多的就是线图了，主要是为了掌握服务器、应用运行的负载趋势。在图形软件展示中，用得最多的就是Grafana。

展开

1

Linuxer
2019-12-11

表格、线图、柱状图各条形图、饼图这些是用得比较多的，我觉得散点图、热图这些可能会比较有用，另外有一个问题大家都是用什么来绘制图形呢?我之前都是用的excel结合宏来生成图表

张智凯

2019-12-11

使用柱状图、饼图、折线图比较多 ，柱状图一般使用是按两个维度拆分，可以很清楚的看出一个变量在另一个变量中的占比，经常一个维度是时间，看某项指标随时间的占比，饼图一般比较某项指标的占比，饼图有时我也会拆分两次。折线图就是看指标随时间的变化趋势了，还可以将数据再拆分，或者将多条折线放到一个图里，感觉折线图比较方便的是可以看到指标数值的变化，柱状图对比比较明显，饼图可以看占比。

展开